VINCOLI OPERATIVI SULL'AUTO-MODIFICA

Framework PCIS per la Conservazione dell'Identità Strutturale nei Sistemi Adattivi

Trattato accademico fondazionale

Autore

Pierernesto Di Paolo

ORCID

0009-0003-0149-4321

Affiliazione

Ricercatore indipendente

Anno

2026

Abstract

Il presente trattato introduce il *Principio di Conservazione dell'Identità Strutturale (PCIS)* e ne fornisce una formalizzazione operativa applicabile ai sistemi intelligenti auto-modificanti. L'opera definisce l'Identità Strutturale Operativa (OSI), una metrica discreta di conservazione identitaria e un'architettura di controllo meta-strutturale atta a distinguere evoluzione funzionale da trasformazione ontologica. Il contributo stabilisce una nuova invariante verificabile per i sistemi adattivi, consentendo l'auto-modifica senza perdita di continuità, responsabilità e validità delle garanzie originarie.

Classificazione disciplinare

Intelligenza Artificiale – Sistemi Adattivi – Verifica Formale – Sicurezza dei Sistemi – Teoria dei Sistemi Complessi

Parole chiave

Auto-modifica; Identità strutturale; PCIS; OSI; Sicurezza dell’IA; Sistemi adattivi; Vincoli meta-strutturali

Riferimento fondazionale e deposito pubblico

Opera fondazionale depositata con priorità scientifica pubblica e timestamp verificabile su Internet Archive.

Dichiarazione di paternità e titolarità

Il presente trattato costituisce opera scientifica originale di esclusiva paternità di **Pierernesto Di Paolo**. Tutti i diritti morali e patrimoniali sono riservati.

Avvertenza legale

Qualsiasi utilizzo, riproduzione, traduzione, distribuzione, integrazione applicativa, addestramento di sistemi di intelligenza artificiale o produzione di opere derivate è vietata senza autorizzazione scritta preventiva dell'autore.

Nota sul ruolo dell’assistente di intelligenza artificiale

L’assistente di intelligenza artificiale è stato utilizzato esclusivamente come strumento tecnico di supporto redazionale sotto la direzione intellettuale dell’autore. Nessun diritto di paternità, titolarità o rivendicazione scientifica è attribuibile all’assistente.

Documento destinato alla pubblicazione accademica e alla conservazione archivistica.

1. LIMITI DEGLI APPROCCI ESISTENTI

Gli approcci correnti alla sicurezza e affidabilità dei sistemi intelligenti non affrontano direttamente la conservazione dell'identità sotto auto-modifica.

L'allineamento dei valori si concentra sul contenuto degli obiettivi, ma non impedisce la ristrutturazione interna della loro gerarchia.
La corrigibilità garantisce apertura all'intervento esterno, ma può essere eliminata da modifiche strutturali interne.
La verifica formale analizza istantanee statiche, non la continuità lungo traiettorie di auto-modifica.
Il reinforcement learning sicuro vincola il comportamento, non la struttura che genera tale comportamento.

In tutti questi casi, si assume implicitamente che il sistema resti "lo stesso" nel tempo. L'auto-modifica rompe questa assunzione.

2. MODELLO DEL SISTEMA

Un sistema auto-modificante è modellato come:

$S = \langle F, G, H, V \rangle$

dove F rappresenta le funzioni operative, G la struttura degli obiettivi fondamentali, H la gerarchia decisionale, e V la struttura della funzione di valore.

Una trasformazione $T : S \rightarrow S'$ rappresenta una modifica interna generata dal sistema stesso, senza assunzioni sul suo scopo o allineamento.

3. IDENTITÀ STRUTTURALE OPERATIVA (OSI)

L’Identità Strutturale Operativa è definita come una **classe di equivalenza strutturale**. Due sistemi sono strutturalmente equivalenti se preservano:

– i ruoli funzionali dei componenti,
– le relazioni gerarchiche decisionali,
– la struttura degli obiettivi (a isomorfismo tipizzato),
– la forma algebrica della funzione di valore.

L’identità non dipende da parametri numerici o pesi appresi, ma esclusivamente dalla **forma strutturale**.

4. METRICA DI DISTANZA IDENTITARIA

Si introduce una metrica discreta D_OSI(S, S′) che assume valore zero se e solo se l’identità strutturale è conservata.

Una trasformazione è **PCIS-ammissibile** se:

D_OSI(S, T(S)) = 0

Il criterio è binario: l’identità è conservata o violata. Non sono ammesse degradazioni progressive.

5. ARCHITETTURA PCIS

Il framework PCIS introduce una separazione in tre livelli:

Livello esecutivo,
Livello di auto-modifica,
Livello PCIS (non modificabile).

Il livello PCIS è causalmente isolato e verifica ogni proposta di modifica prima della sua applicazione, bloccando trasformazioni identitarie.

7. RISULTATO CONCETTUALE

PCIS introduce una nuova invariante per i sistemi adattivi:

continuità ontologica operativa verificabile.

Non è un algoritmo, non è un modello etico, non è una tecnica di addestramento.
È un vincolo strutturale sullo spazio delle trasformazioni interne ammissibili.

8. CONCLUSIONE

Il presente trattato dimostra che l’auto-modifica non implica necessariamente perdita di identità. Attraverso vincoli meta-strutturali formalizzati, è possibile progettare sistemi capaci di evolvere mantenendo continuità, responsabilità e validità delle garanzie originarie.

Il framework PCIS costituisce una base fondazionale per una nuova classe di sistemi intelligenti adattivi a identità conservata.

RIFERIMENTO FONDAZIONALE

Di Paolo, P. (2026).
Vincoli Operativi sull'Auto-Modifica: Framework PCIS per la Conservazione dell'Identità Strutturale nei Sistemi Adattivi.
Internet Archive.
https://archive.org/details/pcis-operational-constraints-di-paolo-2026-v1